Coherence-driven argumentation to norm consensus

Joseph, S.; Prakken, H.

Published in:
Proceedings of the 12th International Conference on Artificial Intelligence and Law, Barcelona, 2009

Coherence-Driven Argumentation to Norm Consensus

Sindhu Joseph
Artificial Intelligence Research Institute, IIIA
Spanish National Research Council, CSIC
Bellaterra (Barcelona), Catalonia, Spain

Henry Prakken
Department of Information and Computing Sciences, Utrecht University, and
Faculty of Law, University of Groningen
The Netherlands

ABSTRACT

In this paper coherence-based models are proposed as an alternative to logic-based BDI and argumentation models for the reasoning of normative agents. A model is provided for how two coherence-based agents can deliberate on how to regulate a domain of interest. First a deductive coherence model presented, in which the coherence values are derived from the deduction relation of an underlying logic; this makes it possible to identify the reasons for why a proposition is accepted or rejected. Then it is shown how coherence-driven agents can generate candidate norms for deliberation, after which a dialogue protocol for such deliberations is proposed. The resulting model is compared to current logic-based argumentation systems for deliberation over action.

Keywords

Deductive coherence, norm deliberation, normative agents, argumentation

1. INTRODUCTION

The research reported in this paper is in the context of a coherence-based approach to the modelling of autonomous artificial agents. One of the fundamental properties that a human mind tries to preserve is its coherence. Any new information is tended to be evaluated for its coherence with the whole before accepting or rejecting. Taking this intuition to artificial systems, a coherence-based agent theory [16, 21] provides the agent with a mechanism to preserve the coherence of its cognitions. With this approach, beliefs, desires or intentions are only accepted if they belong to a coherent whole. That is, a coherence-based agent not only selects the set of actions to be performed, but also looks for the best set of goals to be pursued and beliefs to be accepted, making it a more dynamic model of cognitions.

In contrast, traditional BDI theories [20] do not have such a measure of coherence built into the theory. This means that agents lack the discriminative power to evaluate a cognition, thus making them less autonomous. Further, approaches that extend BDI [6] equate decision making to a process to evaluate actions (intentions) with respect to certain fixed beliefs and goals. However, this makes it hard to prioritise goals or discover potential conflicts. In recent argument-based versions of BDI [3, 4, 2] goals can be prioritized and certain conflicts can be discovered. However, they tend to be more brittle since support and defeat relations between arguments and the acceptability of arguments cannot be a matter of numerical degree, while sets of acceptable arguments cannot contain conflicts. On all these points a coherence approach is meant to provide more flexibility, since in reality support, attack and acceptability are often a matter of degree. One aim of this paper is to introduce coherence models as a more flexible alternative to logic-based argumentation models.

A dynamic model of agency is all the more necessary in normative agents where conflicts between goals, beliefs and external norms are more frequent. A generic coherence-based framework was proposed in Joseph et al. [14], applying the coherence-based approach to normative reasoning of a single agent. They show how an agent driven by its coherence evaluations can decide to adopt norms when it is coherent to do so, and dynamically decide to violate a previously adopted norm when new beliefs makes it less coherent to comply with the norm. However, since [14] only treat a single agent case, they do not further explore the scenario where several agents can deliberate about norms.

In the present paper we address the latter topic by extending this research to a multi-agent setting, in which two coherence-driven agents aim to reach agreement about how to regulate a certain aspect of their society. We aim to define a dialogue protocol for this situation and to model how the individual agents can behave within this protocol. In particular we address the following research questions. Given an agent's beliefs, its private goals (what the agent thinks is good for itself) and its social goals (what the agent thinks is good for the society to which it belongs):

> How can an agent generate norms for discussion on how to promote social goals?
>
> How can an agent decide to accept a norm proposed by another agent given its social and private goals?
>
> How can two agents reach consensus to adopt or discard a norm?

This paper is organised as follows. In Sections 2 and 3 we present the coherence model of [14] and how it is used to model coherence-driven normative agents. In Section 4

ICAIL-2009 Barcelona, Spain

we propose a dialogue system for two-agent deliberation on which norms to adopt. We illustrate our approach with an example in Section 5 and compare it with related research in Section 6. We conclude in Section 7.

2. COHERENCE FRAMEWORK

Since we consider coherence-driven agents, in this section we summarise a generic coherence framework that will allow us to build coherence-based agents. The framework, introduced by Joseph et al [14, 13], is based on Thagard's formulation of the theory of coherence as maximising constraint satisfaction [21]. This theory is based on the assumption that pieces of information can be associated with each other, the association being either positive or negative. The framework of [14, 13] differs from other coherence-based agent theories [6, 16] as it modifies the way an agent framework is perceived by making the associations in the cognitions explicit in representation and analysis. That is, in this framework coherence is treated as a fundamental property of the mind of an agent. Further, it is generic and fully computational. The core notion is that of a *coherence graph* whose nodes represent pieces of information and whose weighted edges represent the degree of coherence or incoherence between nodes. In the following we briefly introduce the necessary definitions of this framework to understand the formulation of coherence-driven norm deliberation.

Definition 1. A *coherence graph* is an edge-weighted undirected graph $g = \langle V, E, \zeta \rangle$, where

1. V is a finite set of nodes representing pieces of information.
2. $E \subseteq \{\{v, w\} | v, w \in V\}$ is a finite set of edges representing the coherence or incoherence between pieces of information, and which we shall call *constraints*.
3. $\zeta : E \rightarrow [-1, 1] \setminus 0$ is an edge-weighted function that assigns a negative or positive value to the coherence between pieces of information, and which we shall call *coherence function*.

Every coherence graph is associated with a number called the *coherence of the graph*. Based on Thagard's formalism, this can be calculated by partitioning the set of nodes V of the graph in two sets, $\mathcal{A}$ and $V \setminus \mathcal{A}$, where $\mathcal{A}$ contains the accepted elements of V, and $V \setminus \mathcal{A}$ contains the rejected ones. The aim is to partition V such that a maximum number of constraints is satisfied, taking their values into account. A constraint is satisfied only if it is positive and both the end nodes are in the same set, or negative and the end nodes are in complementary sets. Formally:

Definition 2. Given a coherence graph $g = \langle V, E, \zeta \rangle$, and a partition $(\mathcal{A}, V \setminus \mathcal{A})$ of V, the *set of satisfied constraints* $C_{\mathcal{A}} \subseteq E$ is given by

$$C_{\mathcal{A}} = \left\{ \{v, w\} \in E \;\middle|\; \begin{array}{l} v \in \mathcal{A} \text{ iff } w \in \mathcal{A} \text{ when } \zeta(\{v, w\}) > 0 \\ v \in \mathcal{A} \text{ iff } w \notin \mathcal{A} \text{ when } \zeta(\{v, w\}) < 0 \end{array} \right\}$$

All other constraints (in $E \setminus C_{\mathcal{A}}$) are said to be *unsatisfied*.

Definition 3. Given a coherence graph $g = \langle V, E, \zeta \rangle$, *the strength of a partition* $(\mathcal{A}, V \setminus \mathcal{A})$ of V is given by

$$\sigma(g, \mathcal{A}) = \frac{\sum_{\{v,w\} \in C_{\mathcal{A}}} | \zeta(\{v, w\}) |}{| E |}$$

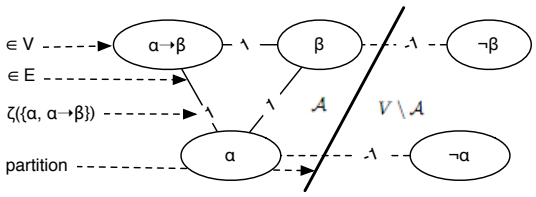

Figure 1: A typical coherence graph with a coherence maximising partition

Notice that by Definitions 2 and 3,

$$\sigma(g, \mathcal{A}) = \sigma(g, V \setminus \mathcal{A}) \quad (1)$$

Definition 4. Given a coherence graph $g = \langle V, E, \zeta \rangle$ and given the strength $\sigma(g, \mathcal{A})$, for all subsets $\mathcal{A}$ of V, *the coherence of* g is given by

$$\kappa(g) = \max_{\mathcal{A} \subseteq V} \sigma(g, \mathcal{A})$$

If for some partition $(\mathcal{A}, V \setminus \mathcal{A})$ of V, the strength of the partition is maximal (i.e., $\kappa(g) = \sigma(g, \mathcal{A})$) then the set $\mathcal{A}$ is called the *accepted* set and $V \setminus \mathcal{A}$ the *rejected* set of the partition. A typical coherence graph is shown in Figure 1.

Due to Equation 1, the accepted set $\mathcal{A}$ is never unique for a coherence graph. Moreover, there could be other partitions that generate the same value for $\kappa(g)$. Here we mention a few criteria to select an accepted set among the alternatives. One such criterion makes use of one of Thagard's principles of deductive coherence[21] (which we will present below), namely, that "intuitively obvious propositions have an acceptability on their own". If $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ are sets from all those partitions that maximise coherence of the graph g, we say an accepted set is the one in which the intuitively obvious propositions belong. Further, the coherence of the sub-graphs $(g|_{\mathcal{A}_i}, i \in [1, n])$ gives us an indication of how strongly connected they are. The higher the coherence, the more preferred the corresponding accepted set. And lastly, an accepted set with more elements should be preferred to another with fewer elements.

We now need a way in which coherence graphs can be constructed. That is, we need to define function ζ. As the nature of the relationship between two pieces of information (corresponding to the different types of coherence such as *explanatory, deductive, conceptual, analogical, perceptual, and deliberative* as specified by Thagard) can vary greatly, we do not have a unique coherence function. Thagard proposes principles to characterise coherence in each of the different types. Here we define one such coherence function, which is inspired by Thagard's principles of *deductive coherence*.

Thagard's principles state that 1) *deductive coherence is a symmetric relation* 2) *a proposition coheres with propositions that are deducible from it,* 3) *propositions that are used together to deduce something cohere with each other,* 4) *the more hypotheses it takes to deduce something, the less the degree of coherence,* 5) *contradictory propositions are incoherent with each other*[1]. Since some of these principles

[1]Here we do not formalise the principle that 6) *intuitively obvious propositions have a degree of acceptability on their own.* As just explained, we use it instead to select among accepted sets.

make sense only in the context of a theory presentation, we assume a theory presentation $\mathcal{T}$ in a multi-valued propositional logic while formalising these principles. We use a multi-valued logic to model uncertainty in agents, though Thagard's principles, we assume, are based on a boolean world. We formalise Thagard's principles in terms of a *coherence function* $\zeta_{\mathcal{T}}$ which extracts a coherence value between two nodes if either one implies the other, or together they are used to imply a third node. We also normalise the values between $[-1, 1]$. Semi-formally $\zeta_{\mathcal{T}}$ is defined as follows (the full details can be found in [14]):

(1) the size of the smallest set of formulas that is needed to make α and β satisfy principle 2 (i.e. such that $\mathcal{T}, \alpha \vdash \beta$ but not $\alpha \vdash \beta$);

(2) the size of the smallest set of formulas that is needed to make α and β satisfy principle 3 (i.e. such that $\mathcal{T}, \alpha, \beta \vdash \gamma$ but not $\alpha, \beta \vdash$);

In both cases it holds that the larger $\mathcal{T}$, the lower the coherence between α and β (principle 4). Contradiction is treated as in case 2 with the contradictory propositions together implying falsehood ($\perp$).

(3) the truth value of β (in case 1) or γ (in case 2): in both cases it holds that the greater the truth value, the higher the coherence between α and β.

In fact, the strength as defined above is separately defined for two directions: from α to β and from β to α (in the latter case α and β are interchanged in condition (1). To obtain the final, symmetric strength (principle 1) between α and β, the highest of the two directional strengths is taken.

There may be a need to find those nodes that support a given node. This is mostly required to defend a coherence-based decision. One of the criticisms raised against coherence-based decision making is the lack of justification behind a decision. To counter this criticism we introduce two simple notions *support set* and *conflict set* of a node.

The intuition is that if two nodes have a positive coherence between them, then they reinforce or give support to each other. However, we cannot take support from rejected nodes as they do not actively take part in the decision making process. And if two nodes have a negative coherence between them, then they counter each other. However, the conflict set of a node should contain the support of those nodes that conflict with the node.

Definition 5. Given a coherence graph $g = \langle V, E, \zeta \rangle$ and given a coherence maximising partition $(\mathcal{A}, V \setminus \mathcal{A})$, *the support set of a node* v is given by

$$S(v) = \{w \in \mathcal{A} | e(\{v, w\}) > 0\} \tag{2}$$

Definition 6. Given a coherence graph $g = \langle V, E, \zeta \rangle$ and given a coherence maximising partition $(\mathcal{A}, V \setminus \mathcal{A})$, *the conflict set of a node* v is given by

$$C(v) = \left\{ \begin{array}{l} w \in V | e(\{v, w\}) < 0 \\ w \in S(w) | e(\{v, w\}) < 0 \end{array} \right\} \tag{3}$$

3. COHERENCE-DRIVEN AGENTS

We now describe the architecture of a coherence-driven agent and in particular how such an agent generates new norms in the context of norm deliberation.

3.1 Agent Architecture

A *coherence-driven agent* acts based on maximisation of coherence. We consider cognitive agents based on the BDI theory [20], using an adaptation of Casali et al. [7] based on multi-context systems (MCS), which incorporate graded cognitions. The grade in a cognition represents the degree to which an agent believes (desires or intends) a particular cognition. We use graded cognitions (represented in a multi-valued logic) to incorporate reasoning under uncertainty into our agent framework. Then, an MCS models the representation and interaction between these graded cognitions.

The MCS specification of an agent contains three basic components: contexts, logics for contexts, and bridge rules between contexts.. Contexts in a multi-context BDI are the contexts of belief, desire, and intention cognitions. The deduction mechanism of MCS is based on two kinds of inference rules, internal rules $\vdash_i$ inside each context and bridge rules B between contexts. Internal rules allow an agent to draw consequences within a context, while bridge rules allow to embed results from one context into another [11]. Thus, an agent is defined as a family of interconnected contexts: $\langle \{C_i\}_{i \in I}, B \rangle$ where

- each context $C_i = \langle L_i, A_i, \vdash_i, T_i \rangle$ consists of a language L_i, a set of axioms A_i, and a deductive relationship $\vdash_i$. They define the logic for the context and its basic behaviour as constrained by the axioms. In addition a theory $T_i \in L_i$ is associated with each context, which represent the particular axioms of the context.

- the set B of bridge rules consists of inference rules with premises and conclusions in different contexts.

We add further structure to [7] by associating a coherence graph to the theory of each of the contexts. Since theories in each contexts are expressed as coherence graphs our bridge rules carry consequences from one graph to another.

3.1.1 Contexts and Bridge Rules

We assume that each agent has beliefs stored in its belief context C_B and goals stored in its desires context C_D (both individual and social goals). No relation is assumed between the C_D contexts of different agents, so they may not only have different individual goals but also have different social goals. The intentions of each agents are in the intention context C_I[2]. We also assume that each agent has a normative context C_O, which stores its opinions on the norms that should hold in the normative institution of which it is part.

Each context has its own language, logic and theories expressed as coherence graphs. The context and coherence graphs vary depending on the logic selected. Here we consider a graded BDI logic based on Lukasiewicz connectives as defined in [7]. The norm logic of the norm context is adapted from the work of Godo et al [9] on probabilistic deontic logic. Beliefs formulas are of the form $(B\varphi, r)$ which states *the confidence on the belief of the propositional formula* φ *is at least* r. Similarly we have desire, intention and norm formulas with modalities D, I and O and where the grades represent preferences. Further, theory $\mathcal{T}_B$ of the belief context is expressed as belief graph $g_B = \langle V_B, E_B, \zeta_{\mathcal{T}_B} \rangle$ where the coherence function $\zeta_{\mathcal{T}_B}$ is based on the deduction relation $\vdash_B$ defined on the belief logic. Theories $\mathcal{T}_D$ and $\mathcal{T}_O$ correspond to the desire and norm contexts.

[2]In this paper, since we focus on generating obligations, we do not concentrate on actions.

3.2 Norm Generation

We next discuss how coherence-driven agents can generate norms which, if obeyed, achieve one or more social goals that the agent thinks are important. (Recall that social goals are what an agent thinks is good for its society while private goals are what the agent thinks is good for itself.)

Conte et al [8] specify certain conditions under which an agent adopts a norm. Among other things, the norm shouuld be be instrumental to solving some of the social goals of the agent. We extend this principle to specify conditions under which a new norm is generated. A new norm, we claim, stems from an unsatisfied social goal and a belief that certain actions under certain conditions (can be empty) can achieve this goal. We express this with the help of a bridge rule that says *if the goal context implies a social goal ψ and the belief context implies a belief $\phi \rightarrow \psi$ then the normative context contains an obligation ϕ.*

$$\frac{C_B : (B(\phi \rightarrow \psi), \alpha), C_D : (D\psi, \beta)}{C_O : (O\phi, f(\alpha, \beta))}$$

(Here $f(\alpha, \beta)$ is a function that computes the grades of a complex formula given the grades of its subformulas.) If applied naively, this bridge rule will result in too many obligations: if there is more than one way to achieve ψ, then all of them will be turned into obligations, which would overconstrain the normative institution: what we want instead is to make only one way to achieve the social goal obligatory, to increase the agents' degree of autonomy. Another aspect not taken into account by this bridge rule is that realising ϕ may frustrate another social goal, i.e., it may hold that $\phi \rightarrow \neg\psi'$ where ψ' is another social goal of the agent.

To deal with these problems, the obvious similarity can be exploited between this bridge rule and the well-known practical syllogism "If I want ψ and ϕ realises ψ, then I should intend to do ϕ". Walton (1996) formulated this as one of his presumptive argument schemes, with as main critical questions "are there other ways to realise ψ?" and "does ϕ also have unwanted consequences?". In recent years several AI researchers have formalised this argument scheme in formal argumentation systems (e.g. [3, 4, 2]). The key idea here is that positive answers to Walton's two critical questions give rise to counterarguments.

Our task is to model the same idea in our coherence approach. As coherence theory is developed to make sense of such contradictions between pieces of information and identify those that cohere most together, modelling the above scenario is natural using this theory. Coherence maximisation partitions the cognitions including the obligations in such a way that the most coherent set of cognitions and obligations is selected. Note that the basic relationship we model here is that between goals and norms, in which different ways to achieve the same goal negatively cohere with each other. However, our framework uses only deduction as the underlying relation, in which the set $\{p \rightarrow g, q \rightarrow g, p, q\}$ is consistent (here p and q are different ways to achieve goal g). Hence we add an additional constraint to make these alternatives inconsistent. That is, for each goal g in an agent's desire context, we consider the set of all implications $p_1 \rightarrow g, \cdots, p_n \rightarrow g$ in its belief context and we add formulas $\neg(Op_i \& Op_j)$ to its norm context for all p_i and p_j such that $1 \leq i < j \leq n$. Then two obligations Op_i and Op_j negatively cohere with each other since they are alternatives.

This method deals with the first of Walton's critical questions of the practical syllogism (are there alternative ways to realise the same goal?). To deal with his second critical question (does ϕ also have unwanted consequences?) a bridge rule is needed that expresses the negative version of the practical syllogism: if the goal context implies a social goal ψ' and the belief context implies a belief $\phi \rightarrow \neg\psi'$ then the normative context contains an obligation $\neg\phi$.

$$\frac{C_B : (B(\phi \rightarrow \neg\psi), \alpha), C_D : (D\psi, \beta)}{C_O : (O\neg\phi, f(\alpha, \beta))}$$

Then, in cases where an action achieves some but frustrates other social goals, our deductive coherence measure makes the obligations that result from the positive and negative version of the practical syllogism negatively cohere.

4. DEFINITION OF PROTOCOL AND RELATED NOTIONS

We next present a protocol for two-agent deliberation about norm proposals. Dialogues are triggered by a set of mutually accepted social goals (the 'focal goals'). Dialogues are about how best to promote the achievement of these goals by enacting norms (in the hope that the agents of the relevant society will obey the norms and thus help realise the desired effects.) During a dialogue additional social goals may be proposed by each agent and, if accepted by the other agent, norms for these additional goals may be proposed, or norm proposals for the focal goals may be evaluated in terms of their effect on the additional social goals. Besides social goals, the agents may also have their own secret private goals. These are not made public during a dialogue but are used internally by the agent that holds them to decide about making or accepting a proposal.

4.1 Main ideas

A general feature of the protocol is that it is for 'theory building' dialogues. Although the agents exchange arguments, the effect of these is that the agents jointly build a coherence graph, which records the coherence relations between the beliefs, goals and norms mentioned in the arguments, and which thus records the joint understanding of a problem. Arguments can contain norm proposals (by applying one of the above bridge rules) or can be about goals or beliefs. An argument's premises and conclusion are added as nodes to the joint coherence graph, together with the newly induced positive and negative constraints. The joint graph is then used by the protocol to define turntaking, relevance of moves and the dialogue outcome, in ways explained below. Besides the joint coherence graph each agent also has its own internal coherence graph (which may also be updated or revised during a dialogue but which remains hidden for the other agent). This graph is used by the agent to make its internal decisions about what to say in the dialogue (e.g. whether to make or accept a certain proposal). An agent's private goals are incorporated into its internal coherence graph.

The reason for choosing a theory-building approach is that in deliberations about promoting the achievement of social goals by enacting norms the public understanding of a problem is crucial: since the goals addressed are social and the norms bind everyone within the relevant society, the reasons for a consensus should ideally be public. This contrasts with

argument-based negotiation [19], where the negiotating parties are self-interested so that all that counts is whether an argument persuades the hearer to do something in the dialogue (like accepting or revising a proposal) that is beneficial to the speaker. In consequence, protocols for argument-based negotiation usually are not of the theory-building kind but define the outcome of a dialogue purely in terms of explicit acceptances and rejections. Some persuasion protocols are also of that kind, which is suitable when the participants' only goal is to win a dispute. However, when public interests are at stake, a theory-building approach arguably is better.

The idea of theory-building dialogues is not new (see e.g. [12]), but in most current dialogue systems for argumentation the theory built during a dialogue is a set of arguments or some related structure (such as a dialectical graph). By contrast, in the present dialogue system the theory built during a dialogue is a coherence graph. In fact, the protocol will require of arguments that when added to the joint coherence graph, there is indeed a positive coherence in the graph between the argument's premises and conclusion. Thus in our system the notion of an argument is not basic but derived: the basic reasoning/inferential structure is not a set of arguments but a coherence graph, and the inferential machinery applied to the joint theory is not an argument-based logic but a coherence calculus.

The protocol enforces relevance and coherence of dialogues in two ways. Initially, a norm proposal must be made for a social goal that triggered the deliberation. Subsequently, each agent must make sure that its position is best satisfied in that in the joint coherence graph its norm proposals are better realised then those of the other agent. The latter is an implicit relevance mechanism: argument moves must be chosen such that they improve the speaker's current position, which implies that they must somehow pertain to what has been said so far. To capture this, at each stage of the dialogue preferred partitions of the joint graph are identified for each player, which are the partitions in which their own norm proposals are best satisfied. Each player must aim to have the most coherent preferred partition.

Another important element of the protocol is that as soon as a player has succeeded in having the most coherent preferred partition, the turn shifts to the other player, who must then try to have the most coherent preferred partition, and so on. This builds a dialectical element into dialogues that promotes the efficient exploration of all sides of a problem (cf. [15]'s 'immediate-response' disputes). A dialogue ends in agreement when both players' preferred partitions accept the same set of norms.

4.2 Definition of the dialogue system

We now formally define the topic and communication languages L_t and L_c and the protocol.

Let L_t consist of the union of the agents' context languages for beliefs, desires, intentions and norms. Then L_c consists of expressions Φ *since* Γ such that Φ and all elements of Γ are well-formed formulas of L_t. (Below such expressions will be called *arguments*; at first sight, this notion of argument would seem to be wildly unconstrained but, as indicated above and formalised below, the protocol enforces relevance in various ways.) A *move* is a pair (p, x) where x is an expression from L_c and p is the player who utters x (sometimes we will abuse notation and refer to x only as a move, leaving the speaker implicit). A *dialogue* is a sequence of moves. For any dialogue $d = m_1, \ldots, m_n, \ldots$ the sequence $m_1, \ldots, m_i$ is denoted by d_i, where d_0 denotes the empty dialogue. For any dialogue d and move m the notation d, m stands for the result of appending m to d, i.e., for d as continued by m.

Definition 7. For any dialogue d the *joint coherence graph* $g(d) = \langle V(d), E(d), \zeta(d)\rangle$ associated with d is defined as follows (we leave the coherence function implicit since it can be deduced from the other elements by the definitions of [14]):

- $V(d_0) = E(d_0) = \emptyset$ while $\zeta(d)$ is undefined;
- For any move $m = \Phi$ *since* Γ :
  - $V(d, m) = V(d) \cup \{\varphi\} \cup \Gamma \cup C$, where:
    - if $m = (O\psi, \alpha)$ *since* $(B(\psi \rightarrow \chi), \beta), (D\chi, \gamma), S$[3] then $C = \{(\neg(O\psi \wedge O\psi'), f(\alpha, \alpha')) \mid d$ contains a move with argument $(O\psi', \alpha')$ *since* $(B(\psi' \rightarrow \chi), \beta'), (D\chi, \gamma), S$ such that $\psi \neq \psi'\}$;
    - otherwise $C = \emptyset$
  - $E(d, m) = \{(v, v') \mid v, v' \in V(d, m)$ and $\zeta_(v, v')$ is defined$\}$

The joint coherence graph is initially empty. Each move adds its premises and conclusion as new nodes, after which the edges and coherence values are recalculated according to the definition of ζ (cf. Section 2). In addition, if a move proposes a norm in alternative to an earlier proposal for the same goal, we also add the corresponding constraint between the two norms as a new node.

Next it is defined how well a dialogue satisfies a player's norm proposals.

Definition 8. A norm $(O\phi, \alpha)$ is *proposed* by player p in dialogue d if d contains a move (p, x) where the conclusion of x is $(O\phi, \alpha)$.

A goal $(D\psi, \gamma)$ is *addressed* by p in d if d contains a move (p, x) where x is of the form $(O\phi, \alpha)$ *since* $(B(\phi \rightarrow \psi), \beta), D\psi, S$ (= applying the first bridge rule of Section 3.2).

A partition $(\mathcal{A}, V \setminus \mathcal{A})$ of $g(d)$ is *potentially preferred* by player p if the accepted set $\mathcal{A}$ of the partition contains a norm proposed by p for each goal addressed by p in d.

A partition $(\mathcal{A}, V \setminus \mathcal{A})$ of $g(d)$ is *preferred* by player p if it is potentially preferred by p and there is no other potentially preferred partition of $g(d)$ by p with a higher coherence value. Let $P_p(d)$ be any partition of $g(d)$ preferred by p.

Definition 9. A dialogue d *best satisfies* player p if the coherence of its preferred partitions of $g(d)$ is higher than the coherence of the preferred partitions of $g(d)$ of its opponent.

Technically, a *protocol* is a function Pr that assigns to any legal dialogue a set of moves which are its legal continuations. A dialogue is *legal* if any move in it is a legal continuation of the sequence to which it is appended. If $Pr(d) = \emptyset$ then d is a *terminated* dialogue.
As explained above, each dialogue is assumed to be initiated by a set $F = \{(D\psi_1, \alpha_1), \ldots, (D\psi_n, \alpha_n)\}$ of *focal social goals*. The protocol is then defined as follows.

Definition 10. For any dialogue d, $m = (p, x) \in Pr(d)$ iff:

[3] S is a, possibly empty, set of additional premises. For exmaples see Section 5 below.

1. p is the player to move in d;
2. if $d = d_0$ then s is of the form $(O\phi, \alpha)$ *since* $(B(\phi \rightarrow \psi), \beta), (D\psi, \gamma), S$ where $(D\psi, \gamma)$ is a focal goal;
3. if $E(d, m)$ contains positive support links from each premise of x to its conclusion;
4. if the coherence value of p's preferred partitions in $g(d, m)$ is not higher than the coherence value of p's preferred partitions in $g(d)$, then
   (a) either m is p's first proposal for a goal addressed in d;
   (b) or m repeats a proposal for a norm by p'.
5. d contains no move (p, x);
6. the players do not agree in d.

Furthermore, we have that player p is to move in d_i if either d_i best satisfies p' or no player is best satisfied in d_i and p' was to move in d_{i-1}.

To comment on these rules, the first rule is obvious while the second rule says that each discussion starts with a proposal for a norm that (if complied with) achieves some focal social goal. Each next move may be an argument of any form, as long as it respects the remaining protocol rules. Rule (3) says that each move must be an argument in that in the resulting joint coherence graph, the premises of the move must positively cohere with its conclusion. Rule (4) says that each move must either improve the position of the speaker, or make the speaker's first norm proposal for a goal addressed in d, or accept a norm proposal by the other party. Rule (5) prevents a player from repeating his own moves, while rule (6) makes sure that a dialogue terminates after the players have reached agreement.

For defining agreement we need the following notation. For any partition $P = (\mathcal{A}, V \setminus \mathcal{A})$ of graph g let $N_p(P)$ denote the norms proposed by p belonging to $\mathcal{A}$.

Definition 11. The players p and p' *agree* in dialogue d if all focal goals have been addressed in d and there exist preferred partitions $P_p(d)$ and $P_{p'}(d)$ of $g(d)$ such that $N_p(P_p(d)) = N_{p'}(P_{p'}(d))$.

In words, the players agree if they have discussed all focal goals and if they have preferred partitions that contain the same set of norms for all goals addressed in the dialogue (which may include more goals than just the focal goals, namely, if a move has proposed a new social goal).

4.3 Internal Deliberation

We now sketch the internal deliberation of each player p to generate and evaluate proposals. Coherence-driven agents make decisions based on coherence maximisation, same is true for the cases of generation and evaluation of proposals.

4.3.1 Generate a New Move

We assume that at any time the coherence graph of an agent is closed under the application of the bridge rules. The accepted set resulting from the coherence maximising partition is the base for generating new moves. Moves are of the form Φ since Γ. Φ can be any element of the accepted set of a coherence maximising partition, and then Γ is the set of support nodes of Φ. Among the possible Φ's, an element is chosen based on its priority. In the case where the deliberation is on norms, norms can be given priority over other elements. Given the composite coherence graph of the agent $g = \langle V, E, \zeta \rangle$, an agent performs the following to generate a new move:

1. For all partitions $(\mathcal{A}_i, V \setminus \mathcal{A}_i)$, $\mathcal{A}_i \subseteq V$ calculate the coherence $\sigma(\varsigma_{g'}, \mathcal{A}_i)$ using Definition 3.
2. Using Definition 3 from Section 2, find a coherence maximising partition $\mathcal{A} = \mathcal{A}_i | \max(\sigma(\varsigma_{g'}, \mathcal{A}_i))$. Note that there may be more than one such partition (preferences can be set based on discussions on Section 2).
3. $\Phi = (a\varphi, \alpha)$ such that $\alpha = \max(r | (a\varphi, r) \in \mathcal{A})$ where $a \in \{B, D, I, O\}$. (In the case of moves about norms, $a = O$ and $\mathcal{A}$ is $V_N|_{\mathcal{A}}$ and $(O\varphi, \alpha) \notin g(d)$ (not a previously proposed norm, g(d)= joint coherence graph).
4. The support set $S(\Phi) = S(a\varphi, \alpha)$.
5. Return the dialogue move $m = \Phi$ since $S(\Phi)$
6. If $\Phi = null$, then m is set to $null$.

4.3.2 Evaluate a move

The internal deliberation of player p is similarly based on coherence maximisation. p introduces the received move into its respective coherence graphs and recalculates the composite coherence graph. Upon maximising coherence, if the elements of the move belong to the accepted set of its coherence maximising partition, it accepts the move. Else it generates the reasons for rejecting a move. Given the proposed move $m = (\Phi, S(\Phi))$, a coherence-driven agent should act as follows:

1. Recompute the composite coherence graph $g = \langle V, E, \zeta \rangle$ with the elements of m using bridge rules.
2. For all partitions $(\mathcal{A}_i, V \setminus \mathcal{A}_i)$, $\mathcal{A}_i \subseteq V$ calculate the coherence $\sigma(\varsigma_{g'}, \mathcal{A}_i)$ using Definition 3.
3. Using Definition 3 from Section 2, finds a coherence maximising partition $\mathcal{A} = \mathcal{A}_i | \max(\sigma(\varsigma_{g'}, \mathcal{A}_i))$.
4. If $\Phi \in \mathcal{A}$ and $S(\Phi) \subseteq \mathcal{A}$, then accept m. Else calculate the *conflict set* $C(\Psi)$ for each Ψ such that $\Psi \in \{\Phi\} \cup S(\Phi)$ and $\Psi \notin \mathcal{A}$.

5. EXAMPLE — NORM DELIBERATION

Now we take a real scenario in which two coherence-based agents discuss certain norms for regulating a discussion forum, especially on how often the participants may reply to each others' contributions. The focal goals of the agents are:

- f = efficiency (discussions should not take too long)
- s = coverage (discussions should cover as much relevant material as possible)
- p = fairness (the participants should be treated fairly compared to each other)
- t = quality of contributions (the participants should be stimulated to write high-quality contributions).

In addition, one of the agents has a secret private goal $u = x$ not become a moderator.
With this background, two of the possible ways to achieve these goals and how far they help achieve each of the focal goals are given below:

1. r: *everyone gets one reply.* This promotes efficiency $(r \rightarrow f)$ and quality of individual contributions ($r \rightarrow t$) but demotes coverage $(r \rightarrow \neg s)$. The reason why this promotes quality of contributions is that with just one possible reply everyone will make it as good as possible, since they will not get a second chance. It has no net effect on fairness since on the one hand everyone gets the same number of replies (which is fair) but on the other hand an expert in the field will get less opportunity to say what he wants to say than a layman (which is unfair).

2. q: everyone may reply as long as allowed by the moderator. This also promotes efficiency $(q \rightarrow f)$ since the moderator can be trusted to keep discussions short. It also promotes fairness $(q \rightarrow p)$ since the moderator can be trusted to give experts more replies than novices. It has no particular effect on coverage or quality of contributions (since judging whether everything has been covered is too difficult for the moderator).

Theory	$\mathcal{A}$	$V \setminus \mathcal{A}$
$\mathcal{T}_N$	$(Oq, 1), (O\neg r, 0.8)$	
$\mathcal{T}_B$	$(B(q \rightarrow f), 1), (B(q \rightarrow p), 0.9)$ $(B(r \rightarrow \neg s), 1)$	
$\mathcal{T}_D$	$(Df, 1), (Dp, 0.9), (Ds, 0.8)$	

Table 1: The initial theory of Agent A

Initially, agent A is aware of the social goals f, p and s ($\mathcal{T}_D$ in Table 1) and has the knowledge that q helps achieve two of the goals namely f and p while r hinders realising a goal s ($\mathcal{T}_B$ in Table 1). Hence A generates the norms $(Oq, 1)$ and $(O\neg r, 0.8)$ ($\mathcal{T}_D$ in Table 1 and Figure 2). Since A so far has no incoherence, nor any other ways of achieving its goals, every element falls in the accepted set of its internal coherence graph. Since Oq is preferred over $O\neg r$, A initiates the deliberation protocol with the proposal for norm $(Oq, 1)$, $d = d_0$ (the dialogue moves are in Table 4).

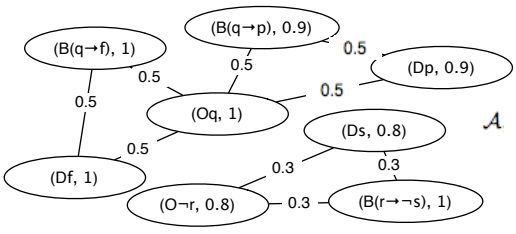

Figure 2: The initial coherence graph of Agent A

Since coherence of the preferred partition of A in the joint graph (Figure 3) is trivially greater than that of agent B, its now B's turn to move. B initially has knowledge of the focal goals f and t and of the facts that r helps achieve these

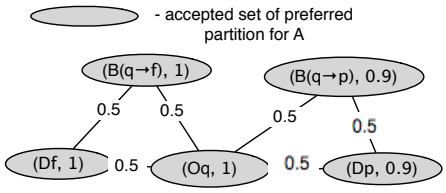

Figure 3: Joint graph after d_0

goals while q hinders achieving one of its secret private goal u (see Table 2). Hence B generates two norms $(Or, 0.9)$ and $(O\neg q, 0.9)$. B also has all the elements in the accepted set of its internal coherence graph so far (in Figure 4 and (Table 2)), as it does not yet know of the conflict between Or and Oq.

Theory	$\mathcal{A}$	$V \setminus \mathcal{A}$
$\mathcal{T}_N$	$(Or, 1), (O\neg q, 1)$	
$\mathcal{T}_B$	$(B(r \rightarrow f), 1), (B(r \rightarrow t), 0.9)$ $(B(q \rightarrow \neg u), 0.9)$	
$\mathcal{T}_D$	$(Df, 1), (Dt, 0.8), (Du, 0.9)$	

Table 2: The initial theory of Agent B

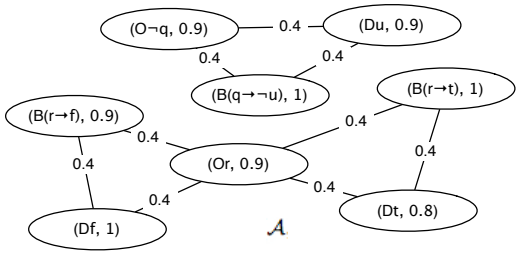

Figure 4: Initial coherence graph of B

After A's move B updates its own coherence graph with A's proposal. However, it is natural to assume that the agents may not have the same preferences on goals. Hence, even though B incorporates the new information into its theory, the degrees of these cognitions vary according to the preference ranking of the goals. The updated theory of B is in Table 3 and the corresponding coherence maximising partition of its own coherence graph is in Figure 6. Since this partition rejects $(Oq, 0.9)$, B makes a counterproposal for the norm $(Or, 0.9)$. This results in the new joint coherence graph of Figure 5.

Since the coherence of the preferred partition of A in this joint graph is not greater than that of agent B, it is now A's turn to move. Since A has no knowledge of B's proposed norm $(Or, 0.9)$ it updates its internal coherence graph with B's proposal. However, A finds out that r upsets its social goal s. The coherence maximisation (Figure 7) hence rejects $(Or, 0.9)$. The joint coherence graph after A has rejected B's proposal (dialogue d_2) is as shown in Figure 8.

Since coherence of the preferred partition of A in the joint graph (Figure 8) is greater than or equal to that of agent B, its now B's turn to move. B incorporates the new informa-

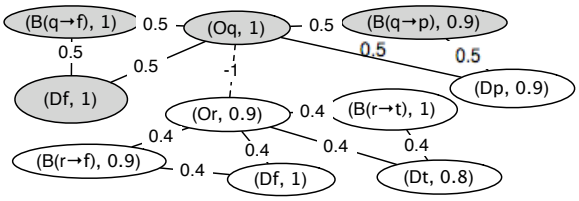

Figure 5: Joint graph after d_1

Theory	$\mathcal{A}$	$V \setminus \mathcal{A}$
$\mathcal{T}_N$	$(Or, 1), (Oq, 1)$	
$\mathcal{T}_B$	$(B(r \rightarrow f), 0.9), (B(r \rightarrow t), 1),$ $(B(q \rightarrow f), 1), (B(q \rightarrow p), 0.9)$	
$\mathcal{T}_D$	$(Df, 1), (Dt, 0.9), (Dp, 0.8)$	

Table 3: Theory of agent B after dialogue d_0

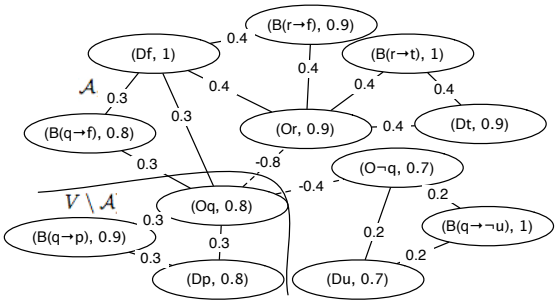

Figure 6: Coherence graph of B after dialogue d_0

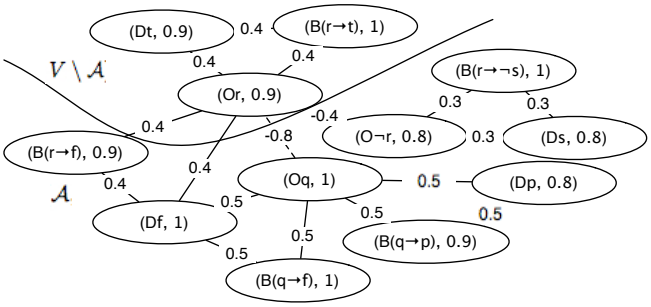

Figure 7: Coherence graph of A after dialogue d_1

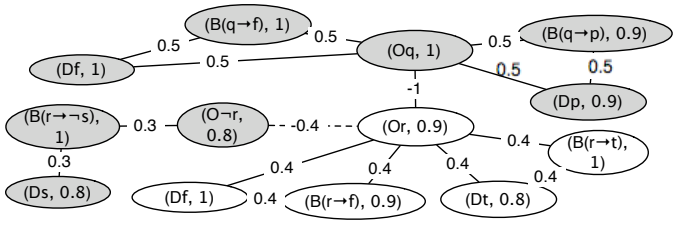

Figure 8: Joint graph after d_2

tion about Or into its theory and calculates the new coherence maximising partition as shown in Figure 9. Due to the fact that the norm $(Or, 0.9)$ upsets social goal s in addition to the competition it has from $(Oq, 1)$, the coherence maximising partition now rejects the norm $(Or, 0.9)$ along with the private goal u, the social goal t and the beliefs relating them. Hence B now proposes the only norm in its accepted set $(Oq, 0.9)$. Now the accepted set of preferred partitions of both A and B in the joint coherence graph (Figure 8) contain the single norm Oq proposed by both of them, so the dialogue ends in agreement. The entire dialogue is as shown in Table 4.

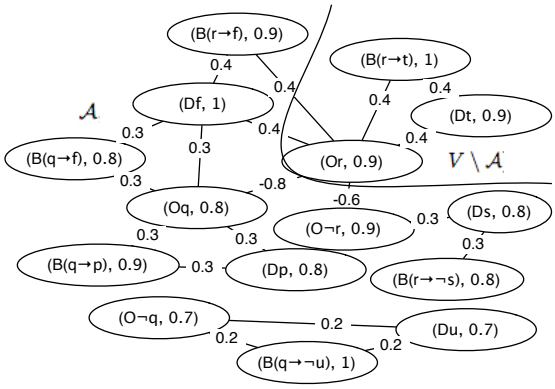

Figure 9: Coherence graph of B after dialogue d_2

Dialogue Id	*Agent*	Φ	Γ
d_0	A	$(Oq, 1)$	$\{(Df, 1), (B(q \rightarrow f), 1),$ $(Dp, 0.9), (B(q \rightarrow p), 0.9)\}$
d_1	B	$(Or, 0.9)$	$\{(B(r \rightarrow f), 0.9), (Df, 1)\}$ $(Dt, 0.9), (B(r \rightarrow t), 1)\}$
d_2	A	$(O\neg r, 0.9)$	$\{((B(r \rightarrow \neg s), 1), (Ds, 0.8)\}$
d_3	B	$(Oq, 0.9)$	$\{(Df, 1), (B(q \rightarrow f), 1),$ $(Dp, 0.9), (B(q \rightarrow p), 0.9)\}$

Table 4: Dialogues between agents A and B

6. RELATED RESEARCH

Coherence models have been earlier applied to legal reasoning by Thagard [22], Amaya [1] and Bench-Capon & Sartor [5]. Thagard and Amaya use explanatory coherence to model scenario-based reasoning about evidence, while Bench-Capon & Sartor use a coherence model in their theory formation approach to case-based reasoning. Thus these proposals model different aspects than ours; moreover, they do not provide protocols for multi-agent deliberation. Further, they do not propose methods to build a coherence graph and hence their models of coherence are not fully computational.

To our knowledge, the first proposal to generate arguments from coherence graphs was by Pasquier [17], However, this model differs from ours in several important respects. Firstly, in [17] coherence is introduced only at the action (intention) level, while in our proposal coherence-driven argumentation can can concern any of the BDI modalities.

Furthermore, [17] provides no definitions for constructing coherence graphs and for generating arguments from them.

We next compare our model to proposals that use logic-based argumentation. We know of no such proposals that address the problems of norm generation and normative agreement. However, norm generation is similar to intention generation by an agent who reasons how to achieve its goals, while normative agreement is similar to reaching agreement on a course of action to solve a problem. For both phenomena logic-based argumentation models have been proposed, so we will compare our model to these.

We must first distinguish between logics and protocols for argumentation. The former define which conclusions can be drawn from a given body of information, while the latter regulate how such a body of information can be constructed in dialogue. Several argument-based logics for intention generation have been proposed. Bench-Capon & Prakken [4] aim to formalise the reasoning model underlying [3]'s dialogue model for disputes over action, while Amgoud & Prade [2] propose an alternative account. The essential ingredient in both approaches consists of two rules for constructing arguments that correspond to our two bridge rules. Bench-Capon & Prakken then apply Prakken's [18] accrual mechanism to aggregate arguments for or against the same intentions, while Amgoud & Prade leave the aggregation of such arguments outside the logic and model it decision-theoretically.

We first note a difference in applying the first bridge rule (the positive practical syllogism), arising from the difference between intentions and norms. While [4] allow to conclude Dr from Dp, $q \Rightarrow p$ and $r \Rightarrow q$ by chaining two applications of the practical syllogism, we do not allow such chaining but only allow to conclude Oq. This is deliberate, since we want to respect the agents' autonomy to decide for themselves how they will comply with the norms they are facing.

The logics of [4, 2] instantiate the general framework of Dung [10], which starts from a set of arguments with a binary defeat relation and then determines which sets of arguments can be accepted together. This is similar to determining partitions of a coherence graph but in approaches that instantiate Dung's format, support and defeat relations between arguments and the acceptability of arguments cannot be a matter of numerical degree, while sets of acceptable arguments cannot contain conflicts. As remarked in the introduction, on all these points a coherence approach is meant to provide more flexibility, since in reality support, attack and acceptability are often a matter of degree. One possible benefit of this is a natural modelling of accrual of arguments for the same conclusion (see e.g. the arguments in the above example). By contrast, in [2] accrual is modelled outside the logic while the logical accrual mechanism of [4] is quite complex. In future research we aim to investigate whether the added flexibility of our coherence approach has other advantages.

On the other hand, a strong point of argument-based approaches is that they yield explicit reasons why an outcome should be adopted or rejected, while coherence-based approaches are often criticised for their lack of transparency. In our approach we have addressed this problem by deriving our coherence measures from the deduction relation of an underlying logic, thus making explicit why two pieces of information are positively or negatively related. This feature was then exploited in our protocol, which contains the notion of an argument.

To compare our protocol with logic-based protocols for reaching agreement over action, the most detailed proposal we know of is that of Atkinson [3], who derives a dialogue protocol from an extended version of Walton's [23] argument scheme for justifying actions and its critical questions. Let us see to what extent our protocol allows her dialogue moves to be moved as arguments in reply to an application of the first bridge rule. Let it be of the form $O\phi$[4] *since* $B(\phi \rightarrow \psi), D\psi$. Note first that we have a restricted domain ontology in that unlike Atkinson we do not distinguish between goals and values, between truth and possibility and between circumstances and actions. All these simplifications are meant to focus on the essence of our proposal, which is its use of the coherence mechanism. These simplifications make that only a number of Atkinson's critical questions are relevant for our model (since we do not distinguish between values and goals, we have replaced Atkinson's term 'value' in CQs 9 and 10 by 'goal'):

- *CQ2: Assuming the circumstances, does the action have the stated consequences?* This can be addressed with an argument for conclusion $B(\neg(\phi \rightarrow \psi))$. This move will introduce a negative coherence link between this conclusion and the original belief $B(\phi \rightarrow \psi)$.
- *CQ5: Are there alternative ways of realising the same consequences?* This can be formulated with an alternative application of our first bridge rule: $O\phi'$ *since* $B(\phi' \rightarrow \psi), D\psi$. Combined with the constraint $\neg(O\phi \wedge O\phi')$ introduced by this move, this move adds a negative support link between $O\phi$ and $O\phi'$.
- *CQ9: Does doing the action have a side effect which demotes some other goal?* We can express this by an application of the second bridge rule. This adds a node $O\neg\phi$ to the joint coherence graph, which negatively coheres with the node $O\phi$.
- *CQ10: Does doing the action promote some other goal?* We can express this by applying the first bridge rule to the other goal, resulting in another argument for the same norm. As shown above, this normally improves the speaker's position and thus naturally models accrual of arguments.
- *CQ11: Does doing the action preclude some other action which would promote some other goal?* This corresponds to the situation that we have $B(\phi \rightarrow \neg\psi)$ and $B(\psi \rightarrow \chi)$ and $D\chi$. Space prevents us from going into logical detail here. Roughly, we can only express this if $\psi \rightarrow \chi$ is necessarily true, i.e., true in all possible worlds: then the argument for $O\phi$ can be countered with an argument for $O\psi$ applying the first bridge rule and further extended to $O\neg\phi$: then $O\phi$ and $O\neg\phi$ negatively cohere in the joint coherence graph.

Concluding, given our restricted domain ontology, our model essentially allows for all argument moves and critical questions proposed by Atkinson; a possible advantage of our approach over Atkinson's is a natural way to model accrual of alternative arguments for the same norm (which is arguably more natural than [4]'s logic-based model of accrual). We leave it for future research to generalise our domain ontology to the full case of Atkinson and to investigate other possible advantages of our approach over hers.

[4]Here the grades are ignored for convenience.

7. CONCLUSION

In this paper we have proposed coherence-based models as an alternative to logic-based BDI and argumentation models for normative reasoning. In particular, we have provided a model for how two coherence-based agents can deliberate to regulate a domain of interest. We first presented a deductive coherence model, in which the coherence values are derived from the deduction relation of an underlying logic; this allowed us to identify the reasons for why a proposition is accepted or rejected. We then incorporated this coherence model in a model of how agents can generate candidate norms for deliberation, after which we proposed a dialogue protocol for such deliberations. The resulting model was shown to be roughly equally expressive as current logic-based deliberation protocols, while it provides a more natural account of accrual of arguments.

In future research we aim to investigate other possible benefits of coherence models over logic-based argumentation models, as well as formal relations between these models. We also want to study properties of our model, such as the conditions under which an agreement is also internally accepted by the agreeing agents. Finally, we aim to extend the expressiveness of our model, for instance by introducing a distinction between goals and values and by using a richer representation language for norms.

8. REFERENCES

[1] A. Amaya. Inference to the best legal explanation. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*. Ashgate Publishing, Aldershot, 2009.

[2] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 34:197–216, 2009.

[3] K. Atkinson. *What Should We Do?: Computational Representation of Persuasive Argument in Practical Reasoning*. PhD Thesis, Department of Computer Science, University of Liverpool, Liverpool, UK, 2005.

[4] T. Bench-Capon and H. Prakken. Justifying actions by accruing arguments. In P. Dunne and T. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 247–258, Amsterdam etc, 2006. IOS Press.

[5] T. Bench-Capon and G. Sartor. A quantitative approach to theory coherence. In B. Verheij, A. Lodder, R. Loui, and A. Muntjewerff, editors, *Legal Knowledge and Information Systems. JURIX 2001: The Fourteenth Annual Conference*, pages 53–62, Amsterdam etc, 2001. IOS Press.

[6] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly Journal*, 2:428–447, 2002.

[7] A. Casali, L. Godo, and C. Sierra. Graded BDI models for agent architectures. In *Computational Logic in Multi-Agent Systems (CLIMA V)*, volume 3487 of *LNAI*, pages 126–143, Berlin/Heidelberg, 2005. Springer.

[8] R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm acceptance. In *ATAL '98: Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages*, volume 1555 of *LNCS*, pages 99–112, Berlin/Heidelberg, 1999. Springer.

[9] P. Dellunde and L. Godo. Introducing grades in deontic logics. In *DEON '08: Proceedings of the 9th international conference on Deontic Logic in Computer Science*, volume 5076 of *LNAI*, pages 248–262, berlin/Heidelberg, 2008. Springer.

[10] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence*, 77:321–357, 1995.

[11] F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics, or: how we can do without modal logics. *Artificial Intelligence*, 65:29–70, 1994.

[12] T. Gordon. The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.

[13] S. Joseph, C. Sierra, and M. Schorlemmer. A coherence based framework for institutional agents. In *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *LNCS*, pages 287–300, Berlin/Heidelberg, 2008. Springer.

[14] S. Joseph, C. Sierra, M. Schorlemmer, and P. Dellunde. Formalising deductive coherence: An application to norm evaluation. In *Normas'08(Extended version), Technical Report(RR-IIIA-2008-02)*, 2009. http://www.iiia.csic.es/ sierra/papers/2009/Coherence.pdf.

[15] R. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.

[16] P. Pasquier and B. Chaib-draa. The cognitive coherence approach for agent communication pragmatics. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 544–551. ACM, 2003.

[17] P. Pasquier, I. Rahwan, F. Dignum, and L. Sonenberg. Argumentation and persuasion in the cognitive coherence theory. In P. Dunne and T. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 223–234, Amsterdam etc, 2006. IOS Press.

[18] H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, New York, 2005. ACM Press.

[19] I. Rahwan, S. Ramchurn, N. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18:343–375, 2003.

[20] A. S. Rao and M. Georgeff. BDI agents: From theory to practice. In *ICMAS–95, First International Conference on Multi-Agent Systems: Proceedings*, pages 312–319, S. Francisco, CA, 1995. MIT Press.

[21] P. Thagard. *Coherence in Thought and Action*. MIT Press, 2002.

[22] P. Thagard. Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, 18:231–249, 2004.

[23] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.